

DOCUMENT RESUME

ED 462 404

TM 033 666

AUTHOR Rosenberg, Reagan; Bryant, Miles
TITLE Discrimination by Young Children across Three Different Survey Response Icons.
PUB DATE 2001-10-00
NOTE 13p.; Paper presented at the Annual Meeting of the American Evaluation Association (St. Louis, MO, November 8-10, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Elementary School Students; Enrichment; Kindergarten; *Kindergarten Children; Primary Education; *Responses; *Summer Programs; Surveys; *Test Construction
IDENTIFIERS Iconic Representation

ABSTRACT

This paper explores the differences in how young children respond to three different types of sample survey response icons. The purpose of the project was to determine if using different types of response icons would result in greater levels of discrimination by children in kindergarten through third grade. The setting was a summer enrichment program in which approximately 400 children enrolled in week-long classes. Surveys were administered at the end of each class. Children were asked to respond to eight items using three different styles of response anchors or icons. In all, 376 children completed surveys. Analysis of the data indicated a significant difference between response icons using "smiley faces" as compared to those using "thumbs-up/down" icons and those using simple "Y" or "N." Children who responded using "smiley faces" did not discriminate between response icons to the degree those did using the response icons in the other two alternative response modes. Program evaluators seeking to gather data from young children should be alerted to the likelihood that the conventional use of "smiley faces" may not reflect the real feelings of these subjects.
(Author/SLD)

Discrimination by Young Children Across Three Different Survey Response Icons

Reagan Rosenberg, MS
Kenexa, Inc. Lincoln Nebraska
rwrosenberg@hotmail.com

Miles Bryant, EdD.
University of Nebraska-Lincoln
Mbryant1@unl.edu

October 2001

Running Head: Response Icons

A paper presented at the annual meeting of the American Evaluation Association, St. Louis,
MO, November 8-10, 2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

R. Rosenberg

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

Abstract

This paper explores the differences in how young children respond to three different types of simple survey response icons. The purpose of the project was to determine if using different types of response icons would result in greater levels of discrimination by children in kindergarten through third grade.

The setting for the study was a summer enrichment program in which approximately 400 children enrolled in week-long classes. Surveys were administered at the end of each class. Children were asked to respond to eight items using three different styles of response anchors or icons. 376 children completed surveys.

Analysis of the data indicated a significant difference between response icons using "smiley faces" as compared to those using "thumbs-up/down" icons and those using simple "Y" or "N." Children who responded using "smiley faces" did not discriminate between response icons to the degree those did using the response icons in the other two alternative response modes.

Program evaluators seeking to gather data from young children should be alert to likelihood that the conventional use of "smiley faces" may not reflect the real feelings of these subjects.

Introduction

This study grew from a program evaluation project of a summer enrichment program in Lincoln, NE. This program provides a great variety of short one-week classes to students in grades kindergarten through ninth grade. The summer enrichment program has been in operation for 15 years.

The evaluation started as a project to quantify information that the summer enrichment program had collected in 2000 over the previous summer. The summer 2000 evaluation used three separate questionnaires, one for parents, one for students from kindergarten through 3rd grade and another for students aged 3rd grade through 9th grade. The kindergarten through 3rd grade questionnaire used closed-ended questions with a three point scale. The three points of the scale consisted of the smiley face, the neutral face and the frown face.

In the data from the summer of 2000 the children in the K-3 courses circled the smiley faces almost exclusively. This overwhelming positive response, and lack of variability in responses led to the research questions that drove this study. Are the children in this program circling the smiley faces because they are genuinely happy with the program? Or, are they circling the smiley faces because they are mostly happy children who have had the opportunity to take part in a summer enrichment program, and because smiley faces are more fun to circle than frowns?

The authors hypothesized that by using other response icons, the evaluation of the program might yield more variability in the children's responses, thus, perhaps providing a better gage of how these younger children perceive their experiences and satisfaction with the program.

Theoretical Basis

Program evaluators who work with young children know that often it is difficult to ask survey questions in ways that produce valid results. This poses a significant problem for programs with young clients. How do the evaluators of these programs gather meaningful data about the perceptions of these young clients. Examples of this problem are plentiful in the annals of research and evaluation.

In a very influential study of young children and the problems associated with assessment, Rosenthal and Jacobson reported achievement scores of young elementary children in attempting to learn about teacher expectations for achievement (Rosenthal and Jacobson, 1968). They subsequently noted that the normed instruments they used were unstable for very young children. This problem is endemic in gathering data from early elementary aged children. Young children may not provide consistent responses nor may they tell evaluators what they really think.

In a typical example, Horowitz, et. al, used a test-retest method to assess the reliability of the Services Assessment for Children and Adolescents (SACA). They used volunteer samples from two different sites consisting of children aged 4 to 17 years old. Their findings show that reliability figures for children aged 9 and 10 were considerably lower for lifetime and 12-month use, and that the younger children's responses suggested that they might have been confused about some questions.

Measurement problems increase as one examines the attitude of even younger children than those in the Horowitz et. al study. In our study, we were gathering data from 5-8 year olds. This is a particularly problematic group.

One of the reasons why survey data gathered from young children can be misleading centers about the problems created by perceptions of social desirability (Weisberg, Krosnick, and Bowen, 1996; Fowler, 1988). All survey respondents exist within a social setting and if the respondent's answer to a survey item is conditioned by what s/he thinks others would find acceptable, the social desirability problem is introduced. A young child may have been indifferent to a particular classroom experience but if she thinks others liked it, she will probably also indicate a liking for the program. And it is certainly possible that the manner in which we ask questions and provide ways to answer those questions have social constructions. A smiley face may well be seen as more socially desirable than a frown or even a neutral face to a young child. When alternative response anchors are utilized such as a thumbs up/?/thumbs down, or a simple Y/?/N social desirability bias may play a lesser role in a young child's response, rendering a more accurate picture of the survey item.

The Study

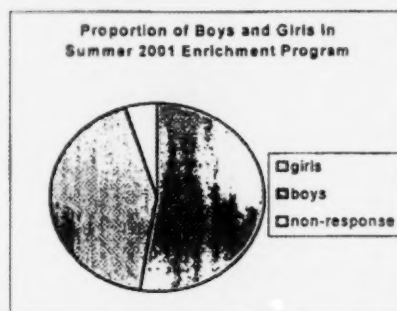
During the summer of 2001 the program offered three enrichment sessions, each lasting one week. An instrument with identical items but different response icons was used to solicit evaluative comments from the students in each of these three different sessions. Along with the smiley face/frown face icons, a thumbs up/thumbs down set, and a Y/N/? set was used. The questionnaires each had the same eight closed ended questions. The ANOVA research design fit well with the structure of this program.

The eight items on the survey were:

1. I liked my (program) class.
2. I wish my class lasted longer.
3. I have a friend in my (program) class.
4. My teacher did a good job teaching.
5. I learned something new in my class.
6. I liked the snacks.
7. This class was fun.
8. I told my parents what I did in class.

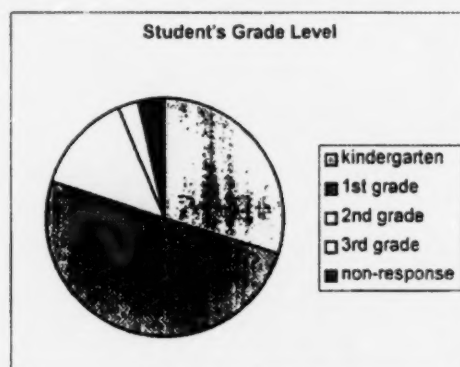
The teachers handed out the instruments in the final hours of each class and were instructed to read the questions to the younger children who were not fully literate. The teachers also collected the instruments as the students completed them. Thus, almost all students who were enrolled were represented in the data of the evaluation. The students were 52.4% girls, 42.8% boys, and 4.8% non-responders. It is possible that the greater number of girls might cause some bias in the results reported later in this study but we did not test for this effect.

Figure 1: Gender Characteristics



The students were also asked to indicate the grade they had completed the previous year. Teachers were asked to help if necessary. 29.8% had been Kindergartners, 50.3% had been first graders, 13.6% had been second graders, 2.8% had been third graders, and 3.7% were non-responders.

Figure 2: Grade Level Characteristics



It is important to note that the preponderance of the students in our study were in the younger groups. Well over 75% had just completed kindergarten or first grade. Additionally, the program directors indicated to us that nearly all of the children were white and came from middle class and above socioeconomic backgrounds. Obviously we could not ask children in these age categories to report family wealth and income data to us.

The summer program ran three separate weeklong sessions. The children in session one were given the smiley face questionnaire (n=121), session two received the Y - N - ? questionnaire (n=148), and session three received the thumbs up/thumbs down questionnaire (n=107).

Session one used:	😊	😐	😞
Session two used:	Y	?	N
Session three used:	👍	?	👎

The responses from the eight items were averaged for each group and the average response was used for the ANOVA.

Results

The responses to surveys using smiley face icons are graphed in figure 3 below. These illustrate positively skewed data produced by responses to the questionnaire using smiley face, neutral face, and frown face icons.

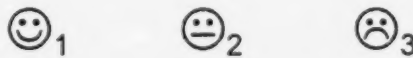
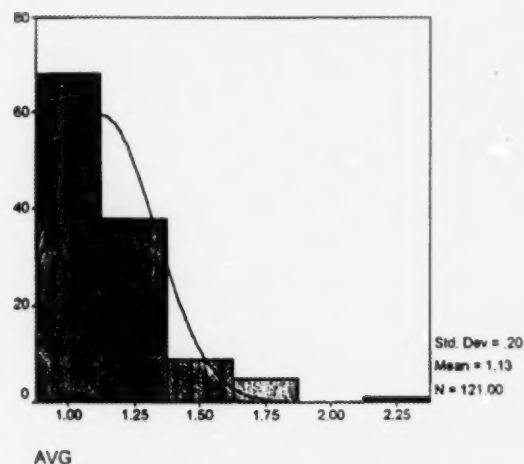


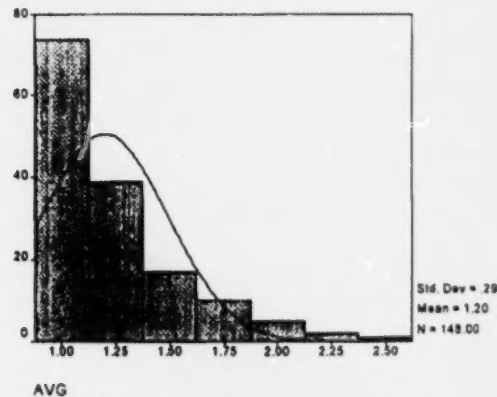
Figure 3: Responses to Smiley Face Icons



The responses to the surveys that asked for Yes, ? or No responses are graphed below. Again, the data indicate a positive skewness with the preponderance of the responses in the most positive category.

Y_1 $?_2$ N_3

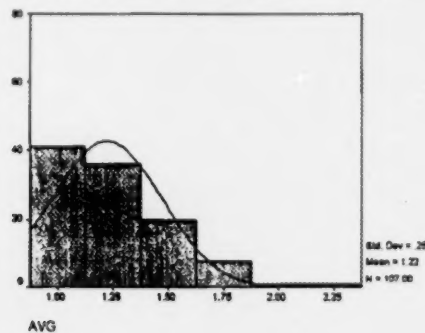
Figure 4: Responses to Y,?,N Icons



Data to response icons of thumbs up/ ? or thumbs down are graphed in Figure 5 below. Though somewhat positively skewed, these data in Figure 5 show a more even distribution. Students in this group appear to discriminate more than do those in the other two groups. Assuming that the groups are similar, this finding appears to support the idea that different response icons might be responsible for different student assessments.

👍₁ ?₂ 👎₃

Figure 5: Responses to thumbs up/thumbs down icons



As Figures 3-5 illustrate, the dependent variables for each group (the response icon alternatives) are all positively skewed. There are three assumptions underlying an ANOVA test; the first is the normal distribution of the dependent variable in each group; the second is homogeneity of variances; and the third is that observations be independent.

We needed to determine if the distribution of our dependent variables violate the assumption of normal distribution of ANOVA. A review by Glass, Peckman, and Sanders (1972) indicated that non-normality has only a slight effect on type I error rate, even for very skewed or kurtotic distributions. With alpha being the probability that the sample mean and the population mean differ, the actual or sample alpha is very close to the nominal or population alpha with respect to the normality distribution. We were comfortable with this assessment of the first problem relative to ANOVA analysis.

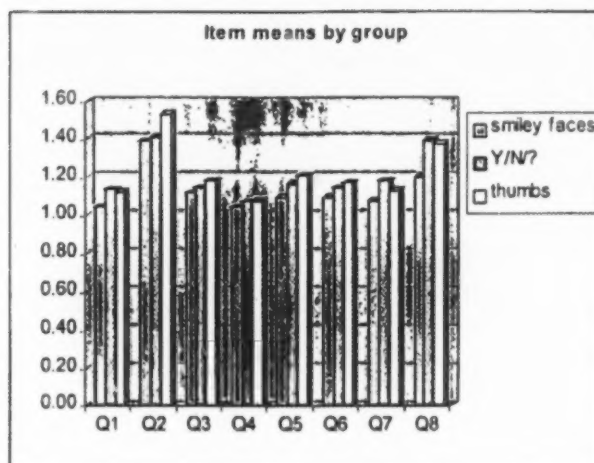
The second assumption of the ANOVA analysis rests upon the homogeneity of population variances. When the group sizes are equal, or approximately equal, the actual alpha stays close to the nominal alpha. Stevens (1999) defined group size as equal if the largest group divided by the smallest group produces a ratio of less than 1.5. The largest group (session 2), $n=148$ divided by the smallest group (session 3), $n=107$ equals 1.38. This ratio of less than 1.5 fulfills the requirement for homogeneity of population variance and we feel our data fall within reasonable parameters relative to this second assumption of ANOVA.

The third assumption of ANOVA is that of the independence of observations. This assumption affects alpha the most. If students filling out the survey instruments speak back and forth, influencing each other, the assumption that each respondent's responses are done without being influence by others is violated. Though one may argue that young students sitting in a class room may have non-independent responses, these student did fill in their own surveys and the questions were geared toward their own experiences in the summer program. We assume although we cannot state with certainty that responses were done independently of each other, students that were supervised by teachers.

For coding purposes, the most positive result was assigned a one, neutral or ? responses were assigned a two and negative responses were assigned a three.

Table 1: Items Means by Group

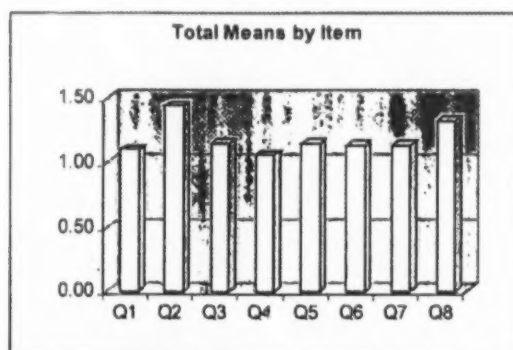
Table 1 below introduces items means for each item across the three groups of respondents. Responses nearest one are most positive, as responses get further from one, they approach the neutral and negative responses.



The highest mean for any of the eight questions asked for feelings about their teacher doing a good job teaching. Most students liked their teachers and felt that they were doing a good job. Responses were lowest, although not low in relative terms, to the question about whether or not the class should have lasted longer. There is not much indication in this data of discrimination by student responders to the survey.

Table 2 below presents the mean scores by individual item. Students agreed most with the two statements about the job their teacher had done and the class being fun

Table 2: Total Means by Item



They were least favorable about wanting class to last longer and the statement that they did inform their parents about what went on in class. Again, these are relative comments. Students were positive about every aspect of their experience.

Table 3 below reports the means and standard deviations across the items by each group. These numerical data show greater variability in some instances than were visible in the graphic presentation of the data.

Table 3: Means and Standard Deviations of Items by Response Icon Group

		Report							
SESSION		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	Mean	1.04	1.39	1.12	1.04	1.09	1.09	1.07	1.20
	N	121	121	121	121	121	121	121	118
	Std. Deviation	.20	.68	.43	.20	.34	.29	.29	.50
2	Mean	1.14	1.41	1.14	1.07	1.16	1.14	1.18	1.39
	N	148	148	148	148	147	147	146	147
	Std. Deviation	.38	.73	.48	.30	.45	.41	.45	.68
3	Mean	1.12	1.53	1.18	1.07	1.21	1.17	1.13	1.37
	N	107	107	106	107	107	107	107	107
	Std. Deviation	.36	.79	.55	.30	.49	.42	.44	.69
Total	Mean	1.10	1.44	1.14	1.06	1.15	1.13	1.13	1.33
	N	376	376	375	376	375	375	374	372
	Std. Deviation	.33	.73	.49	.27	.43	.38	.40	.64

The lower the value, the more positive the score. Thus, one can note that for each item the group using smiley faces (session 1) responded more positively. The responses from the other two groups are also positive, but with greater variation and with uniformly greater standard deviations than that found in the group using smiley faces.

The results of the one-way ANOVA reveal significant differences in response between the three response icon groups beyond the $\alpha=.05$ level.

Table 4: Results of ANOVA Analysis

ANOVA					
AVG					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.580	2	.290	4.521	.011
Within Groups	23.942	373	6.419E-02		
Total	24.522	375			

The ANOVA results allow us to reject the null hypothesis of no difference in treatments. There is a difference in the mean score ratings of the three groups when the opportunity to use different response icons exists.

Post Hoc analysis was performed to analyze our hypothesis that there would be more variation with response icons not using smiley faces. LSD is recommended for three group comparisons when equal variances are assumed.

Table 5: Dependent Variables Analysis

Multiple Comparisons

Dependent Variable: AVG
LSD

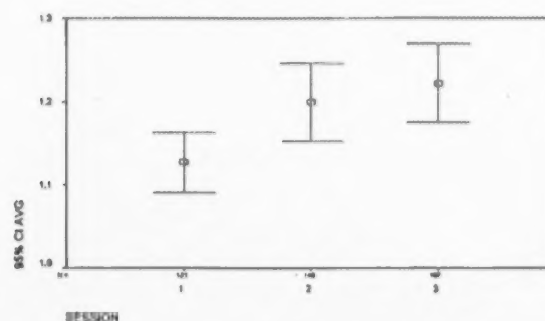
(I) SESSION	(J) SESSION	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
smiley	Y/N/?	-7.2258E-02*	3.105E-02	.020	-.1333	-1.1201E-02
	thumbs	-9.4897E-02*	3.362E-02	.005	-.1610	-2.8787E-02
Y/N/?	smiley	7.225E-02*	3.105E-02	.020	1.120E-02	.1333
	thumbs	-2.2638E-02	3.215E-02	.482	-8.5855E-02	4.058E-02
thumbs	smiley	9.490E-02*	3.362E-02	.005	2.879E-02	.1610
	Y/N/?	2.264E-02	3.215E-02	.482	-4.0578E-02	8.585E-02

*. The mean difference is significant at the .05 level.

The post hoc analysis supports our hypothesis of significant difference between both the smiley and Y?N method of data collection, as well as smiley and thumbs method of data collection. There is a mean difference of .07 between smiley face responses and yes no responses, significant at the .02 level. There is a mean difference of .09 between smiley and thumbs up responses, significant at the .005 level. There is no significant difference between thumbs up and yes/no responses. Another method to analyze differences is using a confidence interval graph.

Below in Figure 6, we provide a visual graphic of confidence intervals. The confidence interval graph shows significant differences between the smiley face and thumb method only. It is between the smiley face response anchor method and thumb response anchor method that there is no overlap between confidence intervals.

Figure 6: Confidence Intervals



Discussion and Conclusions

The LSD post-hoc analysis shows significant differences between session one (smiley face icon) and session two (Y/N/? icon), and between session one (smiley face icon) and session three (thumbs icon). Session one (the smiley faces) yielded almost exclusively smiley faces, the most positive response, with very few other anchors being circled by the young children. While session two and three were also positively skewed, there was a higher prevalence of children circling the question mark, or negative response anchor in these two later sessions. This indicates that using the Y/?/N method as well as the thumbs up/?/thumbs down method may lead to more variability in response by the young children.

When we look at the confidence interval graph, it shows differences between the smiley method and the thumbs method. However, the confidence intervals do not support significant differences between the smiley method and the Y/N method. These results suggest that for future evaluation, evaluators use thumbs response anchors to achieve different responses than smiley face anchors.

The implications of this study should steer those people who are measuring attitudes of young children with smiley faces to consider that different results might be obtained by using different response icons. If so, than one must be continue to be very cautious in interpreting data gathered from young children. Where such data is critical, the authors recommend not using icons with smiley faces upon them.

Citations

Glass, G., Peckman, P., & Sanders, J. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42, 237-288.

Fowler, F.J. (1988). *Survey research methods*. Newberry Park, CA: Sage Publications.

Rosenthal, R. & Jacobson, L. (1968) *Pybmalion in the classroom: Teacher expectations and pupil's intellectual development*. New York: Holt, Rinehart and Winston.

Stevens, J. (1999). *Intermediate Statistics, A Modern Approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Weisberg, H., Krosnick, J. & Bowen, B. (1996) *An Introductin to Survey Research, Polling, and Data Analysis*. Sage Publications, Inc.